

Research Article

Minimum Cost Multiobjective Programming Model for Target Efficiency in Sample Selection

Yousaf Shad Muhammad ¹, Saima Khan,¹ Ijaz Hussain ¹, Alaa Mohamd Shoukry,^{2,3}
Sadaf Shamsuddin,⁴ and Showkat Gani⁴

¹Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan

²Arriyadh Community College, King Saud University, Riyadh, Saudi Arabia

³Workers University, Cairo, Egypt

⁴College of Business Administration, King Saud University, Riyadh, Saudi Arabia

Correspondence should be addressed to Yousaf Shad Muhammad; yousuf@qau.edu.pk

Received 27 September 2018; Revised 19 December 2018; Accepted 3 January 2019; Published 21 February 2019

Academic Editor: Autilia Vitiello

Copyright © 2019 Yousaf Shad Muhammad et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this study, we developed a model which elaborates relationship among efficiency of an estimator and survey cost. This model is based on a multiobjective optimization programming structure. Survey cost and efficiency of related estimator(s) lie in different directions, i.e., if one increases, the other decreases. The model presented in this study computes cost for a desired level of efficiency on various characteristics (goals). The calibrated model minimizes the cost for the compromise optimal sample selection from different strata when characteristic j is subject to achieve at least $(1 - \alpha_j)$ level of efficiency of its estimator. In the first step, the proposed model minimizes the variance for a fixed cost, and it then finds the rise in cost for an α_j percent rise in efficiency of any characteristic j . The resultant model is a multiobjective compromise allocation goal programming model.

1. Introduction

An efficient estimator (estimator with reduced mean square error, MSE) is always desirable while estimating the true population parameter from the sample data. Cost is directly proportional to sample size which increases the efficiency [1] of an estimator as it increases.

This situation becomes more interesting in a multivariate multiobjective study. This estimation mechanism demands a compromise allocation programming model when several goals are competing with each other in a multivariate study. This gives us an inspiration to calibrate a multiobjective compromise allocation model when several estimators are being optimized (minimize mean square error, MSE) on a single feasible space.

Sample allocation in multivariate survey plays an important role in determining the cost and efficiency of an estimator. It has significant impact in case when variance is high among groups. Sampling design, methods of estimation/estimator, and variable cost among groups are main factors

that contribute while allocating sample size. Among the many objectives of a survey, the major two are as follows: it must give efficient estimates for a fixed cost and it should optimize the cost for a complete survey plan with a reasonable efficiency level. In multivariate surveys, sample allocation among heterogeneous strata is an important decision-making technique to increase efficiency of estimates.

Sample selection in survey sampling with a balanced design (i.e., equal allocation) is common. It maximizes the statistical power, but many other scenarios need to address. For example, the groups with large variation usually have to allocate a large sample to reduce variation due to that group. In one-way analysis of variance (ANOVA), if we assume heterogeneous variances among groups, alternate estimators with variable sample size allocation is reasonable to be considered [2]. The heterogeneity among groups and survey cost are the important factors that should be considered while allocating samples. The scarcity of valuable participants of survey is becoming more and more costly, and the survey cost increases.

The mathematical structure of trading between precision or efficiency and cost optimizes the variance or cost function subject to basic constraints. Multiobjective programming models for multiple goals while optimizing transportation problems have been discussed in the study of Roy et al. [3] where authors calibrated these models to minimize transportation cost. The limited budget and the optimal variance suggest only a rough approximation of a sampling design. The efficiency or precision for a fixed cost and vice versa that may actually be achieved by implementation have been discussed in depth, for example, the studies of Cochran [1], Groves [4], and Kish [5]. For an extensive discussion on cost and precision as the only criterion for evaluation, see the study of de Vries [6]. A statistical analysis of the costs and a comprehensive discussion on cost or on error variance trade-off is given in the studies of Kish [5] and Groves [4] for various survey designs.

For an optimum allocation of various characteristics, optimal costs, or optimal variance, the solution of such problems needs some compromise allocation criteria. Varshney et al. [7] discussed in detail the optimal allocation which minimizes some cost vectors, minimizes the weighted average of variances, or maximizes the relative efficiency of the estimators. Many other studies including Bethel [8], Khan et al. [9], Ansari et al. [10], Khan et al. [11], and Varshney et al. [12] worked out different compromise criterion. Most of them solved the allocation problem in a stratified sampling design.

This paper focuses on the trade-off between cost and relative efficiency of estimators. In this context, it investigates the extent to which survey costs and other related components can improve the design and technique for similar such surveys. It tries to take on the problem of reducing the cost of a survey or increasing the efficiency of the estimate. Such analysis are useful in twofold; firstly, it potentially provides a practical solution to the problem in which there is scarcity of information on cost or efficiency of estimates and, secondly, addresses the similarities of sampling designs, survey infrastructure, and population distributions. This information can be obtained from one survey and used in another survey, or it improves the efficiency of same survey estimates. A variety of components are typically a problem, specifically, when some fixed factors are to be compared.

Survey science always recommends a better sampling design, efficient estimator, and reduced total cost of a survey [8–12]. This method comprehensively addresses all the above positive aspects simultaneously and contributes to those studies when researchers are conducting some surveys with competing effects of cost and efficiency of survey. The model minimizes the survey cost for a desired level of efficiency to be achieved for a particular characteristic which is introduced as a constraint in the model. In other words you can say, efficiency is a function of survey cost. Increasing the cost will increase the efficiency of survey. While allocating the cost for some proposed studies, in advance, you will be aware of efficiency level by using some previous studies. Now, if team wants much efficient results, they should expect a relatively high cost for the survey. This model will provide a direct equation of relative change in cost and efficiency even before starting the survey.

Rest of the paper is organized as follows. Section 2 is about problem formulation; in Section 3, we discussed the multiobjective goal programming structure for this model. The successive Sections 4 and 5 elaborate our solution methodology and numerical example, respectively; results and discussion is presented in Section 6, and Section 7 is the summary of our research work.

2. Problem Formulation

Minimum solution is desired either for survey cost or efficient estimates when the remaining solutions are adjusted as constraint. Survey cost and variance or relative efficiency of estimates are major contributing elements of allocation in stratified samples. There are many surveys which are based on a multistage sampling design.

2.1. Sampling Frame. Let we have the data Y_{jhi} (study variable) and X_{jhi} (associated auxiliary variable) for $h = 1, 2, \dots, L$ with $i = 1, 2, \dots, N_h$ sampling units in the h^{th} stratum on $j = 1, 2, \dots, p$ characteristics. Structure of auxiliary attribute can be defined as

$$X_{jhi} = \begin{cases} 1, & \text{if } i^{\text{th}} \text{ unit possesses attribute } X_j \\ & (j = 1, 2, \dots, p) \text{ in the } h^{\text{th}} \text{ stratum,} \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Let \bar{y}_{jhi} and \bar{Y}_{jh} are the sample and population means of the study variable Y_{jhi} , respectively, in the h^{th} stratum for the j^{th} characteristic defined as

$$\begin{aligned} S_{yjh}^2 &= \sum \frac{(y_{jhi} - \bar{Y}_{jh})^2}{N_h - 1}, & \text{where } \bar{Y}_{jh} &= \sum_{i=1}^{N_h} \frac{y_{jhi}}{N_h}, \\ S_{xjh}^2 &= \sum \frac{(x_{jhi} - \bar{X}_{jh})^2}{N_h - 1}, & \text{where } \bar{X}_{jh} &= \sum_{i=1}^{N_h} \frac{x_{jhi}}{N_h}, \\ S_{(yx)jh} &= \sum \frac{(y_{jhi} - \bar{Y}_{jh})(x_{jhi} - \bar{X}_{jh})}{N_h - 1}. \end{aligned} \quad (2)$$

2.2. Survey Cost. Minimizing the survey cost was discussed in the study of Kokan and Khan [13] on various characteristics for a desired precision as a convex problem. Many other authors discussed linear functions which minimize the cost of selecting a sample in stratified sampling design. Suppose a population which has L mutually exclusive strata such that total size is $N = \sum_{h=1}^L N_h$. An independent random sample of size n_h is selected without replacement from each stratum. The linear cost function is

$$C = c_0 + \sum_{h=1}^L c_h n_h, \quad (3)$$

where C is the total budget, c_h ($h = 1, 2, \dots, L$) is per unit cost in the h^{th} stratum, and c_0 is the initial fixed cost of the survey.

Beardwood et al. [14] proposed a cost function, considering the unit and traveling cost as a shortest route among k selected units in the strata, asymptotically proportional to \sqrt{k} , if k is large. A quadratic cost function discussed in the study of Varshney et al. [12] is

$$C = c_0 + \sum_{h=1}^L c_h n_h + \sum_{h=1}^L \tau_h \sqrt{n_h}, \quad (4)$$

where τ_h is per unit travel cost.

A true functional form of cost is important to determine. Practically, in a survey, the unit cost, travel cost, reward to respondent, and labor cost are important factors. A more representative polynomial cost function which includes unit cost, traveling cost, reward to respondent, and labor cost are discussed in the study of Muhammad et al. [15]. They described that labor cost can be computed over time units consumed for a particular respondents. The cost is given as follows:

$$\hat{C} = \sum_{h=1}^L \hat{c}_h n_h + \sum_{h=1}^L \tau_h n_h^\delta + \omega E(T_h), \quad (5)$$

where $\hat{C} = C - c_0$, \hat{c}_h is per unit cost including the reward paid to respondents in the h^{th} stratum, δ is the traveling effect, and τ_h represents the aggregate labor time consumed in h^{th} stratum to collect data on n_h units.

2.3. *Estimators.* Let we have first estimator say $\bar{y}_{j,\text{st}} = \sum_{h=1}^L W_h \bar{y}_{jh}$ of \bar{Y}_j with variance

$$\text{Var}(\bar{y}_{j,\text{st}}) = \sum_{h=1}^L W_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) S_{yjh}^2, \quad (6)$$

where \hat{n}_h and \tilde{n}_h are the two vectors of sample allocation in strata while optimizing $\text{Var}(\bar{y}_{j,\text{st}})$ and $\text{Var}(\bar{y}_{j,\text{lr}})$, respectively. The above two variances are optimum for a same cost, and the only difference between estimates is only due to their structure. The regression estimate partitions an additional component from total variation due to explanatory auxiliary variable involved. If we assume that correlation coefficients among study variable and auxiliary variable are the same or almost the same in all strata, then it is reasonable to suppose that the above two vectors of sample allocation \hat{n}_h and \tilde{n}_h are almost similar. Substituting $\hat{n}_h = \tilde{n}_h = n_h$ in the above equation and solving (ignoring fpc) for efficiency of two estimates, we have

$$\text{EF}(\bar{y}_{j,\text{lr}}, \bar{y}_{j,\text{st}}) = \frac{\text{Var}(\bar{y}_{j,\text{st}})_{\min}}{\text{Var}(\bar{y}_{j,\text{lr}})_{\min}} = 1 + \alpha_j, \quad (11)$$

where $E(\bar{y}_{j,\text{st}}) = \bar{Y}_j$ and $W_h = N_h/N$ is the known stratum weight.

When auxiliary information is given, then the regression estimator can be used. It provides precise estimates when the regression line passes through the origin. Consider the regression estimator for multivariate stratified sampling. The traditional regression estimator is $\bar{y}_{j,\text{lr}} = \sum_{h=1}^L W_h \bar{y}_{j,\text{lr}h}$ where $\bar{y}_{j,\text{lr}h} = \bar{y}_{jh} + b_{jh}(\bar{X}_{jh} - \bar{x}_{jh})$

The variance of $\bar{y}_{j,\text{lr}}$ is given by

$$\text{Var}(\bar{y}_{j,\text{lr}}) = \sum_{h=1}^L W_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) \cdot (S_{yjh}^2 - 2b_{jh}S_{(yx)jh} + b_{jh}^2 S_{xjh}^2), \quad (7)$$

where $b_{jh} = S_{(yx)jh}/S_{xjh}^2$ is the sample regression coefficient in the h^{th} stratum for j^{th} characteristic. The minimum $\text{Var}(\bar{y}_{j,\text{lr}})$ can be computed as [1]

$$\text{Var}(\bar{y}_{j,\text{lr}})_{\min} = \sum_{h=1}^L W_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) S_{yjh}^2 (1 - \rho_{jh}^2), \quad (8)$$

where $\rho_{jh}^2 = S_{(yx)jh}^2 / (S_{yjh}^2 * S_{xjh}^2)$.

The minimum $\text{Var}(\bar{y}_{j,\text{lr}})_{\min}$ can be obtained by minimizing the variance subject to a fixed cost. The difference between two minimum variances (both subject to same fixed cost) is

$$\text{Var}(\bar{y}_{j,\text{st}}) - \text{Var}(\bar{y}_{j,\text{lr}})_{\min}. \quad (9)$$

As discussed in the studies of Cochran [1] and Yousaf and Ijaz [16], the relative increase in optimum variances can be expressed as

$$\frac{\text{Var}(\bar{y}_{j,\text{st}})_{\min} - \text{Var}(\bar{y}_{j,\text{lr}})_{\min}}{\text{Var}(\bar{y}_{j,\text{lr}})_{\min}} = \frac{\sum_{h=1}^L W_h^2 ((1/\hat{n}_h) - (1/N_h)) S_{yjh}^2 - \sum_{h=1}^L W_h^2 ((1/\tilde{n}_h) - (1/N_h)) S_{yjh}^2 (1 - \rho_{jh}^2)}{\sum_{h=1}^L W_h^2 ((1/\tilde{n}_h) - (1/N_h)) S_{yjh}^2 (1 - \rho_{jh}^2)}, \quad (10)$$

where

$$\alpha_j = \frac{\sum_{h=1}^L (W_h^2 S_{yjh}^2 \rho_{jh}^2 / n_{jh})}{\sum_{h=1}^L (W_h^2 S_{yjh}^2 (1 - \rho_{jh}^2) / n_{jh})}. \quad (12)$$

If $k_{jh} = W_h^2 S_{yjh}^2$, the above equation reduces to

$$\alpha_j = \frac{\sum_{h=1}^L (k_{jh} \rho_{jh}^2 / n_{jh})}{\sum_{h=1}^L (k_{jh} (1 - \rho_{jh}^2) / n_{jh})}. \quad (13)$$

3. Multiobjective Goal Programming

3.1. *Single Characteristic Optimization Program.* The optimum allocation within each stratum to minimize the cost for a desired level of efficiency for a particular character “ j ” can

be obtained subject to necessary constraints. The cost function equation (5) minimizes subject to the desired level of efficiency, i.e., $EF(\hat{\theta}_1, \hat{\theta}_2)$, or Eq. (10) along with

$2 \leq n_{jh} \leq N_h$. The optimization program for a minimum cost to the characteristic “ j ” is given:

$$\left. \begin{array}{l} \text{minimize } C_j = f(n_{jh}), \\ \text{subject to } \frac{\sum_{h=1}^L (k_{jh} \rho_{jh}^2 / n_{jh})}{\sum_{h=1}^L (k_{jh} (1 - \rho_{jh}^2) / n_{jh})} \geq \alpha_j, \\ 2 \leq n_{jh} \leq N_h, \\ n_{jh} \text{ are integers and } n_{jh} \in \mathbb{F} \quad \forall h = 1, 2, \dots, L, \end{array} \right\} \text{(Model 1)} \quad (14)$$

where \mathbb{F} is the feasible region established by grid points formed by decision space.

3.2. Multiobjective Goal Program. Our problem is a multivariate problem. We need an allocation within each stratum which should be optimum for all the characteristics. The allocations, optimizing individual characteristic, may differ from one another if characteristics are not correlated. So we

need a suitable Compromise allocation as discussed in the study of Cochran [1].

In the context of data frame discussed above, a total sample of size $n = \sum_{h=1}^L n_h$ is determined that minimizes cost vector for all characteristics $Y_j (j = 1, 2, \dots, p)$. We formulate this problem as a multiobjective, nonlinear integer goal program using the single vector of samples from all strata:

$$\left. \begin{array}{l} C^* = \text{minimize } (C_1, C_2, \dots, C_p), \\ \text{subject to } \frac{\sum_{h=1}^L (k_{jh} \rho_{jh}^2 / n_{jh})}{\sum_{h=1}^L (k_{jh} (1 - \rho_{jh}^2) / n_{jh})} \geq \alpha_j, \\ \text{(or single EF level } \alpha) \\ 2 \leq n_{jh} \leq N_h, \\ n_{jh} \text{ are integers and } n_{jh} \in \mathbb{F} \quad \forall h = 1, 2, \dots, L. \end{array} \right\} \text{(Model 2)} \quad (15)$$

4. Solution Methodology

Our goal program have p goals, which are $j = 1, 2, \dots, p$. To optimize all $j = 1, 2, \dots, p$ goals, we define p vectors of decision variables with each dimension L , i.e., n_{jh} . Decision-maker(s) has the control over these vectors. Among the list of all p goals, every goal has an optimum value say $C_j \forall j = 1, 2, \dots, p$, which is the function of n_{jh} on their fundamentals (14).

There are no assumptions on the generic form discussed in (14) regarding the decision variables. It is the decision-maker(s) who sets a target level for all goals say $C_j^* \forall j = 1, 2, \dots, p$ in (14); they are generally individual optimal values for all objectives, respectively. If any other vector, then the optimal decision vector variables are selected, and the j^{th} goal leads to the following deviation structure:

$$\widehat{C}_j = C_j^* + d_j^+ - d_j^-, \quad (16)$$

where \widehat{C}_j is the j^{th} objective value at the compromised allocation vector and d_j^- and d_j^+ are negative and positive

deviation variables, respectively. d_j^- and d_j^+ are also called unmet goal variables.

If a set of goals is a set of constraints, the decision-maker(s) objective is to optimize each goal. The solution obtained is feasible even if the goal remained unachieved. If hard constraints that are the part of goal programming are violated, the solution turns into infeasible. If $\vec{\mathbb{F}}$ is feasible vector space established by decision vectors, we add the following condition on stratified samples:

$$\vec{n}_j \in \mathbb{F}. \quad (17)$$

The solution thus obtained performs less than our individual optimal solution. We sum up all unwanted deviations as defined in Section 4 and minimize this sum to ensure that our solution is “as close as possible” to our desired goals. Lexicographic goal programming is one solution technique, applied in such problems. It works according to some predefined priority sequence of objectives which are prioritized according to their importance. According to this priority sequence, the first goal is optimized and the remaining goals are compromised by minimizing the

unwanted deviations as discussed in Section 4. A generic form of such compromise allocation is given as

$$\left. \begin{aligned} \text{minimize } \hat{z} &= \left[g_1(n_{1h}), g_2(\underline{d}_2^-, \underline{d}_2^+), \dots, g_p(\underline{d}_p^-, \underline{d}_p^+) \right], \\ \text{subject to } \hat{C}_{jh} + \underline{d}_j^- - \underline{d}_j^+ &\leq \hat{C}_{jh}^*, \\ \hat{n}_j &\in \mathbb{F}, \\ \hat{n}_{jh} \text{ are integers } &\forall h = 1, 2, \dots, L \text{ and } j = 1, 2, \dots, p, \end{aligned} \right\} \text{(Model 3)} \quad (18)$$

where g_1, g_2, \dots, g_p is a sequence of priority of functions and \underline{d}_j^+ and \underline{d}_j^- are directional deviations vectors for maximization or minimization which are unwanted. Weighted goal programming (WGP) technique minimizes a weighted deviation objective function composed by unwanted deviations. Few other techniques are also used in such compromise allocation problems. A goal programming function can be considered as a utility function which always subjects to maximize. It can be described in either form, linear or nonlinear, and sets an aspiration goal within the feasible region [17]. The programming techniques discussed in the studies of Charnes and Cooper [18] and Romero [17] minimize a weighted unmet aspiration $\forall j = 1, 2, \dots, p$ vector of goals in the feasible region. Now, if the goal objective is influenced differently by a negative deviational variable and a positive deviational variable, let \vec{W}_{1j} and \vec{W}_{2j}

are weight vectors of negative and positive deviations on the j^{th} goal, respectively, then above formulation (18) is changed by the following objective [17]:

$$\text{minimize } \hat{z} = \sum_{j=1}^p \left[g_j(\vec{W}_{1j}\underline{d}_j^-, \vec{W}_{2j}\underline{d}_j^+) \right]. \quad (19)$$

If there is some difference between the desired aspiration level and the maximum utility, the Archimedean goal programming interprets it as “the maximization of a separable and additive utility function in the p attributes considered” (Romero [17]). The Minimax (Chebyshev) minimizes the maximum deviation of utility function. This is discussed in the studies of Tamiz et al. [19] and Romero [17]. If D is the maximum deviation from the respective utility or aspiration level, the above program (18) turns into the following equation:

$$\left. \begin{aligned} \text{minimize } D, \\ \text{subject to } \left[\underline{d}_j^-, \underline{d}_j^+ \right] &(\leq \text{ or } \geq) D, \\ \hat{C}_{jh} + \underline{d}_j^- - \underline{d}_j^+ &\leq \hat{C}_{jh}^*, \\ \hat{n}_j &\in \mathbb{F}, \\ \hat{n}_{jh} \text{ are integers } &\forall h = 1, 2, \dots, L \text{ and } j = 1, 2, \dots, p. \end{aligned} \right\} \text{(Model 4)} \quad (20)$$

If φ represents the importance attached to maximizing the achieved aspiration level (20) and $(1 - \varphi)$ weight attached the weighted sum (Eq. (19)) of unwanted deviation model,

then a generalized goal programming model of utility maximization (20) can be written as

$$\left. \begin{aligned} \text{minimize } \hat{z} &= \varphi D + (1 - \varphi) \sum_{j=1}^p \left[g_j(\vec{W}_{1j}\underline{d}_j^-, \vec{W}_{2j}\underline{d}_j^+) \right], \\ \text{subject to } \left[\vec{W}_{1j}\underline{d}_j^-, \vec{W}_{2j}\underline{d}_j^+ \right] &(\leq \text{ or } \geq) D, \\ \hat{C}_{jh} + \underline{d}_j^- - \underline{d}_j^+ &(\leq \text{ or } \geq) \hat{C}_{jh}^*, \\ \hat{n}_j &\in \mathbb{F}, \\ \hat{n}_{jh} \text{ are integers } &\forall h = 1, 2, \dots, L \text{ and } j = 1, 2, \dots, p. \end{aligned} \right\} \text{(Model 5)} \quad (21)$$

The feasible region can be expanded by relaxing the constraint into

$$\varphi \left[\vec{W}_{1j}\underline{d}_j^-, \vec{W}_{2j}\underline{d}_j^+ \right] (\leq \text{ or } \geq) D, \quad (22)$$

and $0 \leq \varphi \leq 1$. D is an arbitrary choice how someone’s positive and negative directional goals may vary. It depends upon how close the modelers or surveyors want to get their results. This strategy provides the leverage to setting the

TABLE 1: Data summary.

h	N_h	W_h	S_{y1h}^2	S_{y2h}^2	S_{x1h}^2	S_{x2h}^2	$S_{(yx)1h}$	$S_{(yx)2h}$	ρ_{1h}^2	ρ_{2h}^2	k_{1h}	k_{2h}	\hat{c}_h	τ_h
1	51	0.6	114.293	21.396	0.238	2.238	1.120	0.483	0.299	0.246	41.1415	7.7026	210	18
2	47	0.4	89.214	19.948	0.255	2.048	0.778	0.700	0.199	0.119	14.2742	3.1917	155	17

feasible region. This will help us to extend our feasible region and allocating an integer sample size. Integer nonlinear programming problems have always a smaller feasible solution grid.

5. Numerical Illustration

We used the data for numerical illustration given in the study of Shafi [20]. We are using these data as our objective to compute trade-off of various estimators over each other how they show change (in sample allocation and cost) against each other for a certain change in the desired level of their MSE.

To explain the data, the study variables are given as follows:

Y_1 = biological yield of wheat varieties, and Y_2 = harvest index of wheat varieties.

Other variables used as covariates are as follows:

X_1 = the number of plants of wheat varieties, and X_2 = total Grain weight of wheat varieties. \bar{Y}_1 is 45.19618523 and \bar{Y}_2 is 16.1223179.

According to protein contents, data are divided into 2 strata. X_{jhi} is the artificial dichotomous variable, the cutoff for quantitative variable to be transformed into attribute is set as the respective stratum mean for each characteristic in the h^{th} stratum. Table 1 shows the summary statistics of the used data.

In this illustration, we suppose that \hat{c}_j is unit cost selection in the h^{th} stratum. A unit of time cost ω and the constant δ is determined using methodology discussed in the studies of Winston [21] and Taha [22]. In the cost function (Eq. (5)), the time taken, to collect the data within h^{th} stratum as well as the whole sample $n = \sum_{h=1}^L n_h$, is considered as the collection of independently and identically distributed random variables, i.e., the time consumed for a unit. All of them follow exponential distribution, and their sum follows a Gamma distribution with parameters say n_h (sample size in h^{th} stratum) with an average of λ_h (Ross [23] and Hogg et al. [24]). Furthermore, the same is true for whole sample from all strata, i.e., $n = \sum_{h=1}^L n_h$, with average $\bar{\lambda}_{st} = \sum_{h=1}^L w_h \lambda_h$, where $w_h = n_h/n$. The total expected time can be computed as $\sum_{h=1}^L E(T_h) = \sum_{h=1}^L \left(\int_0^{\infty} \lambda \exp^{-\lambda t} \sum_{h=1}^L n_h^{-1} / (n_h - 1)! dt \right) = \sum_{h=1}^L (n_h / \lambda)$, [23, 24], where $\lambda = 1/\bar{\lambda}_{st}$. A complete formulation of the problem is shown in Appendix.

6. Results and Discussion

A complete population selected as the sample provides the highest efficiency level of the estimators (variance is minimum). The efficiency is calculated as (see Eq. (10))

$$EF(\bar{y}_{j,lr}, \bar{y}_{j,st}) = \frac{\text{Var}(\bar{y}_{j,st})_{\min}}{\text{Var}(\bar{y}_{j,lr})_{\min}}. \quad (23)$$

For an arbitrary survey cost of 25000 (Figure 1), it is 1.2915 for characteristic one and 1.5645 for characteristic two. We compute minimum cost (21) for varying levels of $\alpha_1 = 0.001, 0.002, \dots, 0.400$ and similarly $\alpha_2 = 0.001, 0.002, \dots, 0.600$ as shown in Figures 1 and 2. Setting α_1 and α_2 beyond these values, solution repeated or model gives no integer solution. A minimum 25000 cost produces 29.15 percent efficient solution for characteristic one and 56.45 for characteristic two, when only single characteristic's variance is optimized, respectively. Different combinations of α_1 and α_2 give a minimum cost solution using compromise allocation structure discussed in (21). We see that compromise solution even reduces our initial survey cost of 25000 if choice of pair of α_1 and α_2 is small. Keeping one of α_1 and α_2 constants and increasing the other increase our cost for a compromise solution. Increasing both parameters simultaneously increase our survey cost (Figure 2). Figures 1(a) and 1(b) are plots of percentage rise in efficiency of an estimator (x -axis) and rise in cost (y -axis) for each characteristic, which shows that if we increase α_1 one percent successively, rise in the compromise cost is not as higher as in the case of second characteristic. Figures 1(c) and 1(d) show sample allocation in each strata for various pair of choice of α_1 or α_2 . This is because each time we make a change in our choice of α_1 or α_2 or both; (21) makes such a selection of samples considering cost and variance of both stratum independently. This might be due to difference in cost and variance structure in both strata. In the above goal programming structure, we can make any percentage to proportionally rise in the variances of both characteristics. For some particular choice of α_1 or α_2 , an integer solution might not be possible as it always returns a grid point of n_1 and n_2 . We restricted the goal programming model to select at least 2 sampling units from strata 1 and 2. Figure 2 is a 3-D plot of percentage rise in efficiency of an estimator 1 (α_1), percentage rise in efficiency of an estimator 2 (α_2), and compromised cost which (21) suggests.

7. Summary

It may not be possible to obtain a compromise integer solution for all possible choices of α_1 and α_2 , i.e., the percentage rise in the variances of characteristics. However, a real value solution may exist for all choices. A high unit cost in a particular stratum may limit its selection. This optimization structure works as a dual procedure in which the cost of the survey is optimized when we require a certain level of efficiency among different estimates, and initially we optimize a compromise variance for a supposed

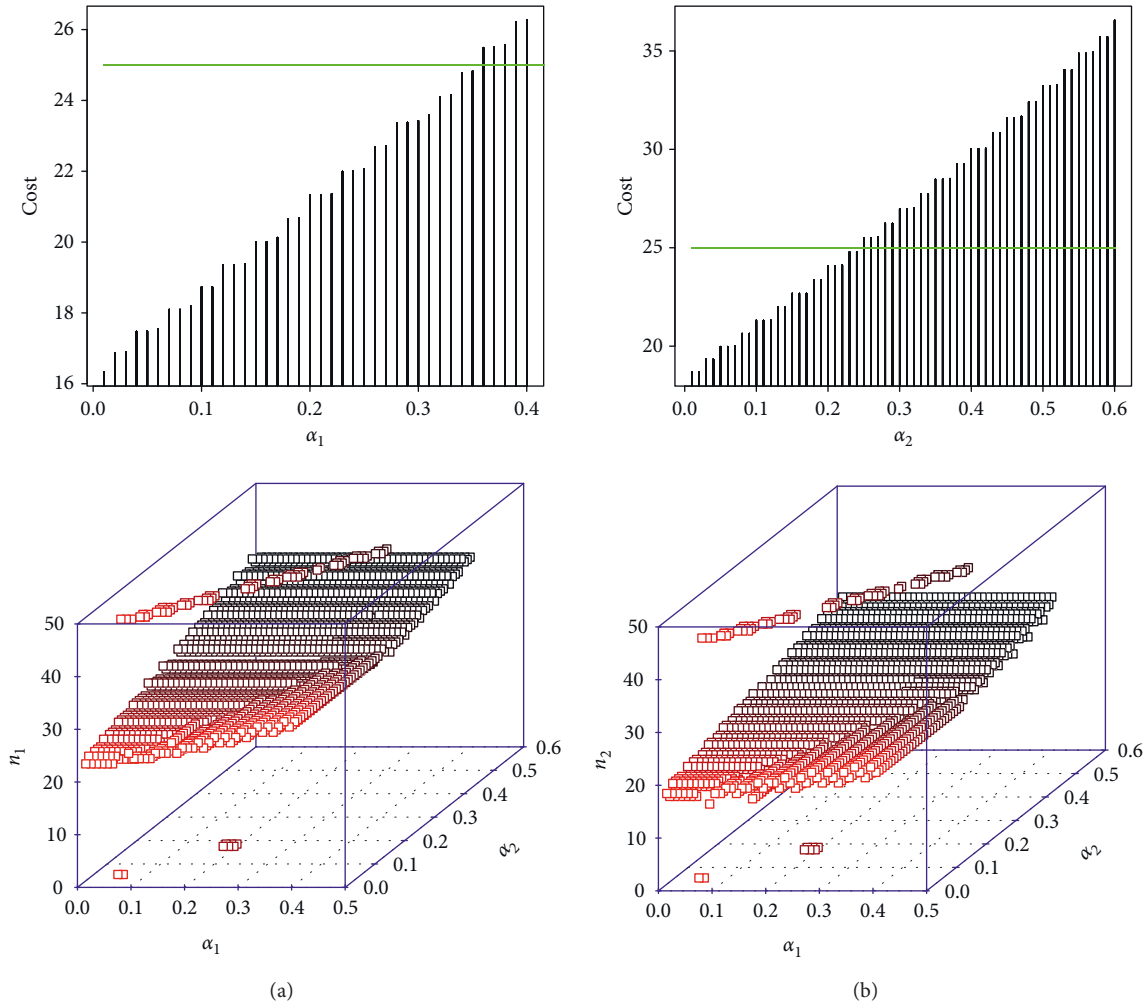


FIGURE 1: Sample allocation and cost for different choices α_1 and α_2 .

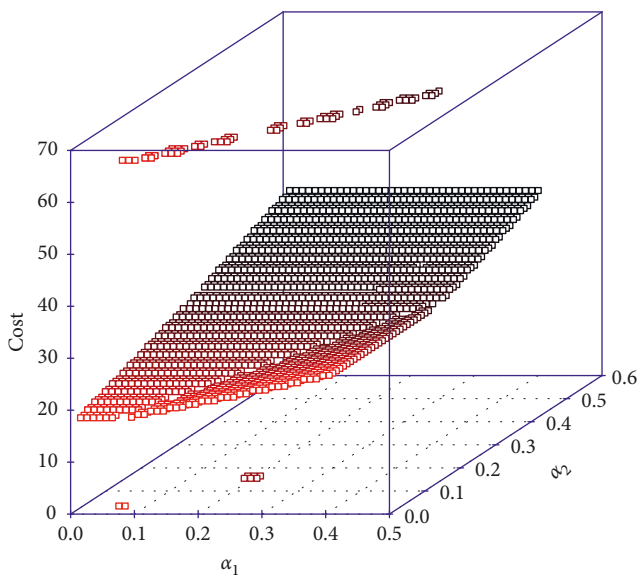


FIGURE 2: Compromise cost (in thousand) for different choices of α_1 and α_2 .

survey cost. Application of other goal programming techniques may also be interesting. We used the nonlinear cost function, but selection of some other cost function can produce different results. Furthermore, we can extend this work using stochastic, dynamic, or stochastic dynamic programming.

Appendix

A. Single Characteristic Models for $(j = 1, 2)$

Our decision variables are $n_{jh} (\forall h = 1, 2, \dots, L)$. Values of \hat{C}_h and τ_h are replaced by Table 1 ($\forall h = 1, 2, \dots, L$). δ replaced with arbitrary values 0.5 and 2.0, and ω is the cost for a unit time of labor (say 100, 150, etc.). Estimation of the Gamma function is replaced with its total expected time $\bar{\lambda}_{st}$ (say 15 minutes, 20 minutes, etc.). The first stage programs for $j = 1$ and $j = 2$ are given below (14). The values of k_{jh} and ρ_{jh} are computed from Table 1:

$$\left. \begin{aligned}
\min \quad & \hat{C}_1 = [210 \ 155] * [n_{11} \ n_{12}]^T + [18 \ 17] * [n_{11}^\delta \ n_{12}^\delta]^T + \omega \sum_{h=1}^L n_{1h} \bar{\lambda}_{1h}, \\
\text{subject to} \quad & \frac{(0.299 * 41.1415/n_{11}) + (0.199 * 14.2742/n_{12})}{(0.701 * 41.1415/n_{11}) + (0.801 * 14.2742/n_{12})} \geq \alpha_1, \\
& 2 \leq n_{1h} \leq N_h, \\
& n_{1h} \text{ are integers and } n_{1h} \in \mathbb{F}, \quad \forall h = 1, 2, \dots, L, \\
\min \quad & \hat{C}_2 = [210 \ 155] * [n_{11} \ n_{12}]^T + [n_{11}^\delta \ n_{12}^\delta]^T + \omega \sum_{h=1}^L n_{1h} \bar{\lambda}_{2h}, \\
\text{subject to} \quad & \frac{(0.246 * 7.7026/n_{21}) + (0.199 * 3.1917/n_{22})}{(0.754 * 7.7026/n_{21}) + (0.881 * 3.1917/n_{22})} \geq \alpha_2, \\
& 2 \leq n_{2h} \leq N_h, \\
& n_{2h} \text{ are integers and } n_{2h} \in \mathbb{F} \quad \forall h = 1, 2, \dots, L.
\end{aligned} \right\} \quad (\text{A.1})$$

B. Multiobjective Model for ($j = 1, 2$)

From the above models, we obtain optimal values, say \hat{C}_1^* and \hat{C}_2^* for two characteristics $j = 1$ and $j = 2$. Selecting an arbitrary value(s) of φ (say 0.1, 0.2, 0.3, ..., 1.0), we establish the following model (let $\varphi = 0.3$, an arbitrary choice):

$$\left. \begin{aligned}
\text{minimize} \quad & 0.3 D + 0.7 (d_1^+ + d_2^+), \\
\text{subject to} \quad & d_1^+ \leq D, \\
& d_2^+ \leq D, \\
& [210 \ 155] * [n_{11} \ n_{12}]^T + [18 \ 17] * [n_{11}^\delta \ n_{12}^\delta]^T + \omega \bar{\lambda}_{st} \sum_{h=1}^L n_h - d_1^+ \leq \hat{C}_1^*, \\
& [210 \ 155] * [n_{11} \ n_{12}]^T + [18 \ 17] * [n_{11}^\delta \ n_{12}^\delta]^T + \omega \bar{\lambda}_{st} \sum_{h=1}^L n_h - d_2^+ \leq \hat{C}_2^*, \\
& \frac{(0.299 * 41.1415/n_{11}) + (0.199 * 14.2742/n_{12})}{(0.701 * 41.1415/n_{11}) + (0.801 * 14.2742/n_{12})} \geq \alpha_1, \\
& \frac{(0.246 * 7.7026/n_{21}) + (0.199 * 3.1917/n_{22})}{(0.754 * 7.7026/n_{21}) + (0.881 * 3.1917/n_{22})} \geq \alpha_2, \\
& 2 \leq n_h \leq N_h, \\
& n_h \text{ are integers and } n_h \in \mathbb{F} \quad \forall h = 1, 2, \dots, L.
\end{aligned} \right\} \quad (\text{B.1})$$

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

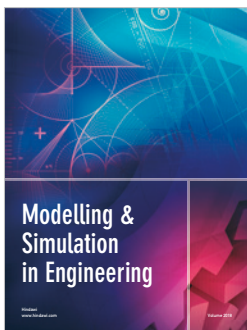
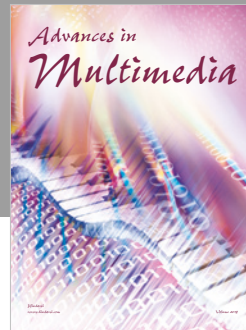
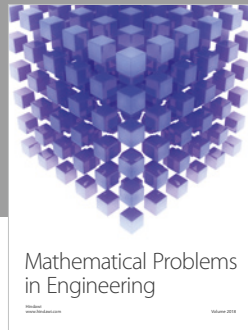
The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Saud University for funding this work through research group no RG-1437-027.

References

- [1] W. G. Cochran, *Sampling Techniques*, John Wiley, New York, NY, USA, 1977.
- [2] R. R. Wilcox, "Adjusting for unequal variances when comparing means in one-way and two-way fixed effects ANOVA models," *Journal of Educational Statistics*, vol. 14, no. 3, pp. 269–278, 1989.
- [3] S. K. Roy, G. Maity, and G.-W. Weber, "Multi-objective two-stage grey transportation problem using utility function with goals," *Central European Journal of Operations Research*, vol. 25, no. 2, pp. 417–439, 2016.
- [4] R. M. Groves, *Survey Errors and Survey Costs*, John Wiley & Sons, Hoboken, NJ, USA, 1989.
- [5] L. Kish, "Optima and proxima in linear sample designs," *Journal of the Royal Statistical Society. Series A (General)*, vol. 139, no. 1, pp. 80–95, 1976.
- [6] W. de Vries, "Are we measuring up...? Questions on the performance of national statistical systems," *International Statistical Review*, vol. 67, no. 1, pp. 63–77, 1999.
- [7] R. Varshney, M. J. Ahsan, and M. G. M. Khan, "An optimum multivariate stratified sampling design with nonresponse: a lexicographic goal programming approach," *Journal of Mathematical Modelling and Algorithms*, vol. 10, no. 4, pp. 393–405, 2011.
- [8] J. Bethel, "An optimum allocation algorithm for multivariate surveys," in *Proceedings of the Survey Research Section*, pp. 209–212, American Statistical Association, Las Vegas, NV, 1985.
- [9] E. A. Khan, M. G. M. Khan, and M. J. Ahsan, "Optimum stratification: a mathematical programming approach," *Calcutta Statistical Association Bulletin*, vol. 52, no. 1–4, pp. 323–334, 2016.
- [10] A. H. Ansari, Najmussehar, and M. J. Ahsan, "On multiple response stratified random sampling design," *Journal of Statistical Mechanics*, vol. 1, no. 1, pp. 45–54, 2009.
- [11] M. G. M. Khan, T. Maiti, and M. J. Ahsan, "An optimal multivariate stratified sampling design using auxiliary information, an integer solution using goal programming approach," *Journal of Official Statistics*, vol. 26, no. 4, pp. 695–708, 2010.
- [12] R. Varshney, Najmussehar, and M. J. Ahsan, "Estimation of more than one parameters in stratified sampling with fixed budget," *Mathematical Methods of Operations Research*, vol. 75, no. 2, pp. 185–197, 2012.
- [13] A. R. Kokan and S. Khan, "Optimum allocation in multivariate surveys: an analytical solution," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 29, no. 1, pp. 115–125, 2018.
- [14] J. Beardwood, J. H. Halton, and J. M. Hammersley, "The shortest path through many points," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 55, no. 4, pp. 299–327, 2008.
- [15] Y. S. Muhammad, J. Shabbir, I. Husain, and M. Abdel-Moemen, "Multi-objective compromise allocation in multivariate stratified sampling using extended lexicographic goal programming with Gamma Cost Function," *Journal of Mathematical Modelling and Algorithms in Operations Research*, vol. 14, no. 3, pp. 255–265, 2014.
- [16] Y. S. Muhammad and I. Husain, "Tradeoff between cost and variance for a multiobjective compromise allocation in stratified random sampling," *Communications in Statistics*, vol. 46, no. 6, pp. 2655–2666, 2016.
- [17] C. Romero, "Extended lexicographic goal programming: a unifying approach," *Omega*, vol. 29, no. 1, pp. 63–71, 2001.
- [18] A. Charnes and W. W. Cooper, *Management Models and Industrial Applications of Linear Programming*, John Wiley & Sons, New York, NY, USA, 1961.
- [19] M. Tamiz, D. Jones, and C. Romero, "Goal programming for decision making: an overview of the current state-of-the-art," *European Journal of Operational Research*, vol. 111, no. 3, pp. 569–581, 1998.
- [20] M. B. Shafi, "Study of physiological and morphological characteristics of wheat for genomic selection," Retrieved from Dissertations and Theses QAU database, Call No. 3541, 2013.
- [21] W. L. Winston, *Operations Research: Application and Algorithm*, Cengage Learning, Boston, MA, USA, 4th edition, 2003.
- [22] H. A. Taha, *Operations Research: An Introduction*, Prentice-Hall, Upper Saddle River, NJ, USA, 9th edition, 2010.
- [23] S. M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists*, Associated Press, New York, NY, USA, 4th edition, 2009.
- [24] R. V. Hogg, J. W. McKean, and A. Craig, *Introduction to Mathematical Statistics*, Pearson, London, UK, 7th edition, 2012.



Hindawi

Submit your manuscripts at
www.hindawi.com

